

Decompositions of rational functions over real and complex numbers and a question about invariant curves

Peter Müller

February 26, 2015

We consider the connection of functional decompositions of rational functions over the real and complex numbers, and a question about curves on a Riemann sphere which are invariant under a rational function.

1 Introduction

Let $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ be the Riemann sphere, and $\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. A *circle* in $\hat{\mathbb{C}}$ is either a usual circle in \mathbb{C} , or a line in $\hat{\mathbb{C}}$. So the circles in \mathbb{C} are just the curves $\Gamma = \lambda(\hat{\mathbb{R}})$, where $\lambda(z) = \frac{az+b}{cz+d} \in \mathbb{C}(z)$ is a linear fractional function (with $ad - bc \neq 0$).

Long ago Fatou suggested to study (Jordan) curves $\Gamma \subset \hat{\mathbb{C}}$ which are invariant under a rational function of degree ≥ 2 . See [3] for recent progress on this. The case that Γ lies in a circle $\lambda(\hat{\mathbb{R}})$ is not interesting, because any rational function $r = \lambda^{-1} \circ s \circ \lambda$ with $s \in \mathbb{R}(z)$ leaves Γ invariant, and there are no other rational functions with this property.

Motivated by his results on invariant curves in [3], Alexandre Eremenko suggested to investigate the following source of invariant curves, and raised two questions about this family:

Question 1.1. *Let $f, g \in \mathbb{C}(z)$ be non-constant rational functions, such that $f(g(z)) \in \mathbb{R}(z)$, so the curve $\Gamma = g(\hat{R})$ is invariant under $r = g \circ f$. Assume that Γ is not contained in a circle.*

(a) *Is it possible that Γ is a Jordan curve? ([3], [2])*

(b) *Is it possible that $r : \Gamma \rightarrow \Gamma$ is injective? ([5], and special case of [4])*

Note that $\Gamma = g(\hat{R})$ is contained in a circle if and only if there is a linear fractional function $\lambda \in \mathbb{C}(z)$ such that $\tilde{g} = \lambda \circ g \in \mathbb{R}(z)$. In this case $f \circ g = \tilde{f} \circ \tilde{g}$ with

$\tilde{f} = f \circ \lambda^{-1} \in \mathbb{R}(z)$. So the decomposition of $f \circ g$ over \mathbb{C} essentially arises from a decomposition over \mathbb{R} .

There are rational functions $f \circ g \in \mathbb{R}(z)$ whose decompositions do not come from a decomposition over the reals. On the other hand, it is known that decompositions of real polynomials over the complex numbers always arise from real decompositions. See Section 5 for more about this.

The purpose of this paper is to give a positive answer to question (a), and a negative answer to a slight weakening of (b). More precisely, regarding (a), we show:

Theorem 1.2. *For every odd prime ℓ there are rational functions $f, g \in \mathbb{C}(z)$ both of degree ℓ , such that*

- (a) $f(g(z)) \in \mathbb{R}(z)$.
- (b) $g : \hat{\mathbb{R}} \rightarrow \hat{\mathbb{C}}$ is injective, so $g(\hat{\mathbb{R}})$ is a Jordan curve.
- (c) $g(\hat{\mathbb{R}})$ is not a circle.

In order to formulate the next two results, we define a weakening of injectivity of rational functions on \mathbb{R} .

Definition 1.3. A rational function $g \in \mathbb{R}(z)$ is said to be *weakly injective on \mathbb{R}* , if there exists $z_0 \in \mathbb{R}$ which is not a critical point of g , and besides z_0 there is no $y_0 \in \hat{\mathbb{R}}$ with $g(z_0) = g(y_0)$.

A partial answer to question (b) is

Theorem 1.4. *Let $f, g \in \mathbb{C}(z)$ be non-constant rational functions, such that $f \circ g \in \mathbb{R}(z)$. Assume that g is weakly injective, and that the curve $\Gamma = g(\hat{\mathbb{R}})$ is not contained in a circle. Then the map $g \circ f : \Gamma \rightarrow \Gamma$ is not injective.*

A slight variant of this theorem shows that for a fairly large class of rational functions from $\mathbb{R}(z)$, each decomposition over \mathbb{C} arises from a decomposition over \mathbb{R} .

Theorem 1.5. *Let $f, g \in \mathbb{C}(z)$ be non-constant rational functions such that $h(z) = f(g(z)) \in \mathbb{R}(z)$ is weakly injective. Then there is a linear fractional function $\lambda \in \mathbb{C}(z)$ such that $\lambda \circ g \in \mathbb{R}(z)$.*

The main ingredient (besides Galois theory) in the proof of the previous two theorems is the following group-theoretic result. (See Section 3 for the notation.)

Proposition 1.6. *Let G be a transitive group of permutations of the finite set Ω . Let σ be a permutation of Ω of order 2 which fixes exactly one element ω , and which normalizes G , that is $G^\sigma = G$. Let G_ω be the stabilizer of ω in G . Then $M^\sigma = M$ for each group M with $G_\omega \leq M \leq G$.*

The proof of Theorem 1.2 uses elliptic curves. The construction was motivated by a group-theoretic analysis similar to the one which led to the proofs of Theorems 1.4 and 1.5.

Acknowledgment. I thank Alexandre Eremenko for inspiring discussions about invariant (Jordan) curves, and Theo Grundhöfer for a careful reading of the somewhat intricate proof of Proposition 1.6, and his suggestion which simplified its final step. Furthermore, I thank Mike Zieve for discussions about an example in the final section and for pointing out its connection with work by Avanzi and Zannier.

2 Non-circle Jordan curves invariant under a rational function

In this section we work out our sketch from [10] and prove Theorem 1.2.

Let E be an elliptic curve given by a Weierstrass equation $Y^2 = X^3 + aX + b$ with $a, b \in \mathbb{R}$. By $E(\mathbb{C})$ and $E(\mathbb{R})$ we denote the complex and real points of E . For $p \in E(\mathbb{C})$ we let \bar{p} be the complex conjugate of p . We use the structure of $E(\mathbb{C})$ as an abelian group, with neutral element 0_E the unique point at infinity.

For general facts about elliptic curves see e.g. [11].

Lemma 2.1. *Let $\ell \geq 3$ be a prime. Then there is a point $c \in E(\mathbb{C})$ of order ℓ , with $\bar{c} \notin \langle c \rangle$.*

Proof. Let $E[\ell] \subset E(\mathbb{C})$ be the group of ℓ -torsion points. Then $E[\ell]$ is isomorphic to the vector space \mathbb{F}_ℓ^2 , and the complex conjugation acts linearly on this space.

Suppose that the claim does not hold, so the complex conjugation fixes each 1-dimensional subspace of $E[\ell]$ setwise. Then the complex conjugation acts as a scalar map. Therefore either $E[\ell] \subset E(\mathbb{R})$, or $\bar{c} = -c$ for each $c \in E[\ell]$. In the latter case write $c = (u, v)$. So u is real and v is purely imaginary. Thus, upon replacing E with the twisted curve $-Y^2 = X^3 + aX + b$ (which is isomorphic over \mathbb{R} to $Y^2 = X^3 + aX - b$), we obtain in either case an elliptic curve E with $E[\ell] \subseteq E(\mathbb{R})$. On the other hand, $E(\mathbb{R})$ is isomorphic to \mathbb{R}/\mathbb{Z} or to $\mathbb{R}/\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ (see e.g. [11, V. Cor. 2.3.1]). However, $\mathbb{R}/\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ does not have a subgroup isomorphic to $\mathbb{Z}/\ell\mathbb{Z} \times \mathbb{Z}/\ell\mathbb{Z}$. This proves the claim. \square

Lemma 2.2. *Suppose that $X^3 + aX + b$ has three distinct real roots. Then there are elements $w \in E(\mathbb{R})$ such that there is no $\hat{w} \in E(\mathbb{R})$ with $w = 2\hat{w}$.*

Proof. If $X^3 + aX + b$ has three distinct real roots, then $E(\mathbb{R}) \cong \mathbb{R}/\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, so any w corresponding to $(s, 1)$, $s \in \mathbb{R}/\mathbb{Z}$ arbitrary, has the property that there is no $\hat{w} \in E(\mathbb{R})$ with $w = 2\hat{w}$. (In this case, $E(\mathbb{R})$ has two connected components, and for each $\hat{w} \in E(\mathbb{R})$ the element $2\hat{w}$ is on the connected component of 0_E .) \square

By an *automorphism* of an elliptic curve we mean a birational map of the curve to itself which need not fix the neutral element.

Pick $c \in E(\mathbb{C})$ of order ℓ such that $\bar{c} \notin \langle c \rangle$, and set $C = \langle c \rangle$.

Let $\Phi : E \rightarrow E' = E/C$ be the isogeny with kernel C . Let $\Phi' : E' \rightarrow E$ be the dual isogeny. Then $\Phi' \circ \Phi : E \rightarrow E$ is the multiplication by ℓ map on E .

For $w \in E(\mathbb{R})$ as in the previous lemma define involutory automorphisms

- β of E by $\beta(p) = w - p$,
- β' of E' by $\beta'(p') = \Phi(w) - p'$, and
- β'' of E by $\beta''(p'') = \Phi'(\Phi(w)) - p'' = \ell w - p''$.

Note that $\beta'(\Phi(p)) = \Phi(w) - \Phi(p) = \Phi(w - p) = \Phi(\beta(p))$, so

$$\beta' \circ \Phi = \Phi \circ \beta \text{ and likewise } \beta'' \circ \Phi' = \Phi' \circ \beta'. \quad (1)$$

In the following we show that

- $E/\langle \beta \rangle$, $E'/\langle \beta' \rangle$, and $E/\langle \beta'' \rangle$ are projective lines, and
- that there are degree 2 branched covering maps ψ , ψ' , and ψ'' from the elliptic curves to these lines,
- such that ψ and ψ'' are defined over \mathbb{R} , and
- that after selecting uniformizing elements of the projective lines, there are unique rational functions f and g such that the following diagram commutes:

$$\begin{array}{ccccc}
 & & \text{multiplication by } \ell & & \\
 & \searrow & & \nearrow & \\
 E & \xrightarrow{\Phi} & E' & \xrightarrow{\Phi'} & E \\
 \downarrow \Psi & & \downarrow \Psi' & & \downarrow \Psi'' \\
 E/\langle \beta \rangle & \xrightarrow{g} & E'/\langle \beta' \rangle & \xrightarrow{f} & E/\langle \beta'' \rangle
 \end{array}$$

We give an algebraic rather than a geometric description of the functions f and g . This has the advantage that the method can be used to compute explicit examples, as we do at the end of this section.

Let $\mathbb{C}(E)$ and $\mathbb{C}(E')$ be the function fields of E and E' , respectively. Let x and y be the coordinate functions with $x(p) = u$ and $y(p) = v$ for $p = (u, v) \in E(\mathbb{C})$. So $E(\mathbb{C}) = \mathbb{C}(x, y)$ with $y^2 = x^3 + ax + b$. The comorphism β^* is an automorphism of order 2 of the real function field $\mathbb{R}(E)$ (recall that $w \in E(\mathbb{R})$). We compute the fixed field of β^* in $\mathbb{R}(E)$: Write $w = (w_x, w_y)$, and set $z = \frac{w_y + y}{w_x - x}$ (this choice of z is taken from [8]). The addition formula for elliptic curves shows that

$$\beta^*(x) = z^2 - w_x - x \text{ and } \beta^*(y) = z(w_x - \beta^*(x)) - w_y.$$

From that we get

$$\beta^*(z) = \frac{w_y + \beta^*(y)}{w_x - \beta^*(x)} = \frac{w_y + z(w_x - \beta^*(x)) - w_y}{w_x - \beta^*(x)} = z,$$

so z is in the fixed field of β^* . Clearly $\mathbb{R}(x, z) = \mathbb{R}(y, z) = \mathbb{R}(x, y)$. Let $F \subseteq \mathbb{R}(E)$ be the fixed field of β^* . From $z \in F$ and $[\mathbb{R}(E) : F] = 2$ we get $F = \mathbb{R}(z)$ once we know that $[\mathbb{R}(E) : \mathbb{R}(z)] \leq 3$. But this holds, as

$$(z(w_x - x) - w_y)^2 = y^2 = x^3 + ax + b,$$

so x has at most degree 3 over $\mathbb{R}(z)$. Now ψ is just the rational function $E \rightarrow \hat{\mathbb{C}}$ for which $z = \psi(x, y)$.

Suppose (without loss of generality) that E' has also a Weierstrass form $Y^2 = X^3 + a'X + b'$, and let x' and y' be the associated coordinate functions. Note that $\Phi((u, v)) = (A(u), B(u)v)$ for rational functions $A, B \in \mathbb{C}(z)$ and all $p = (u, v) \in E(\mathbb{C})$. Therefore $\Phi^*(x') = A(x)$, where $\Phi^* : \mathbb{C}(E') \rightarrow \mathbb{C}(E)$ is the comorphism of Φ .

Pick z' in $\mathbb{C}(x', y')$ such that $\mathbb{C}(z')$ is the fixed field of β'^* . (Here z' can not be taken from $\mathbb{R}(x', y')$, because E' is not defined over \mathbb{R} .) As before, let ψ' be the rational function with $z' = \psi'(x', y')$.

From (1) we obtain $\Phi^* \circ \beta'^* = \beta^* \circ \Phi^*$, so $\Phi^*(z') = \beta^*(\Phi^*(z'))$ and hence $\Phi^*(z') \in \mathbb{C}(z)$. So $\Phi^*(z') = g(z)$ for a rational function $g \in \mathbb{C}(z)$. This is just the algebraic description of g from above. Similarly one computes f .

For the rest of this section, we work with the morphisms rather than the comorphisms. Recall that Ψ and Ψ'' are defined over \mathbb{R} . So $\overline{\Psi(p)} = \Psi(\bar{p})$ for all $p \in E(\mathbb{C})$.

We now prove the required properties of f and g . The assertion about the degrees follows from well-known facts about isogenies.

- (a) As the multiplication by ℓ map is defined over \mathbb{R} , and so are Ψ and Ψ'' , we have $f \circ g \in \mathbb{R}(x)$.
- (b) We next show that g is injective on $\hat{\mathbb{R}}$. Suppose that there are distinct $z_1, z_2 \in \hat{\mathbb{R}}$ such that $g(z_1) = g(z_2)$. Pick $p, q \in E(\mathbb{C})$ such that $\Psi(p) = z_1, \Psi(q) = z_2$. Then

$$\Psi'(\Phi(p)) = g(\Psi(p)) = g(z_1) = g(z_2) = g(\Psi(q)) = \Psi'(\Phi(q)),$$

so $\Phi(p) = \Phi(q)$ or $\Phi(p) = \Phi(w) - \Phi(q)$. Upon possibly replacing q with $w - q$ we may and do assume $\Phi(p) = \Phi(q)$, hence $p - q \in C$.

Recall that Ψ is defined over \mathbb{R} and $\Psi(p) = z_1$ is real. So $\Psi(\bar{p}) = \Psi(p)$, and therefore $\bar{p} = p$ or $\bar{p} = w - p$. Likewise $\bar{q} = q$ or $\bar{q} = w - q$. Recall that $p - q \in C$, and that $C \cap \bar{C} = \{0_E\}$ by the choice of C . So we can't have $(\bar{p}, \bar{q}) = (p, q)$, nor $(\bar{p}, \bar{q}) = (w - p, w - q)$.

Thus, without loss of generality, $\bar{p} = p$ and $\bar{q} = w - q$. So $p \in E(\mathbb{R})$. Note that $p - q$ and $\bar{p} - \bar{q} = p - w + q$ both have order ℓ . Set $r = (p - q) + (\bar{p} - \bar{q}) = 2p - w$. Then $\ell r = 0_E$, and $r \in E(\mathbb{R})$. We obtain $w = 2(p + \frac{\ell-1}{2}r)$ with $p + \frac{\ell-1}{2}r \in E(\mathbb{R})$, contrary to the choice of w .

- (c) Finally, we need to show that $g(\hat{\mathbb{R}})$ is not a circle. Suppose otherwise. Let λ be a linear fractional function which maps this circle to $\hat{\mathbb{R}}$. Then $\lambda \circ g$ maps \mathbb{R} to $\hat{\mathbb{R}}$, so $\lambda \circ g \in \mathbb{R}(x)$.

Then $\lambda \circ \Psi' \circ \Phi = \lambda \circ g \circ \Psi$ is defined over \mathbb{R} , so $\lambda(\Psi'(\Phi(p))) = \lambda(\Psi'(\Phi(\bar{p})))$ for all $p \in E(\mathbb{C})$. As λ is bijective, Ψ' respects β' , and C is the kernel of Φ , we get that for each $p \in E(\mathbb{C})$ either $p - \bar{p} \in C$, or $p + \bar{p} - w \in C$.

In the first case note that $\bar{p} - p \in \bar{C}$, so also $p - \bar{p} = -(\bar{p} - p) \in \bar{C}$, and therefore $p - \bar{p} \in C \cap \bar{C} = \{0_E\}$. So $p \in E(\mathbb{R})$ if the first case happens.

We see that $p + \bar{p} - w \in C$ whenever $p \in E(\mathbb{C}) \setminus E(\mathbb{R})$. Recall that $w \in E(\mathbb{R})$. So $p + \bar{p} - w \in C \cap \bar{C} = \{0_E\}$, and therefore $p + \bar{p} = w$ for all $p \in E(\mathbb{C}) \setminus E(\mathbb{R})$. As $(p+q) + \overline{p+q} = 2w \neq w$ for all $p, q \in E(\mathbb{C}) \setminus E(\mathbb{R})$, we get the absurd consequence that $p + q \in E(\mathbb{R})$ whenever $p, q \in E(\mathbb{C}) \setminus E(\mathbb{R})$, so $E(\mathbb{R})$ is a subgroup of index 2 in $E(\mathbb{C})$. This final contradiction proves all the properties about the functions f and g .

Remark 2.3. For fixed curves E, E' and isogeny Φ as in the proof of Theorem 1.2, and $w \in E(\mathbb{R})$ (which need not fulfill the property of Lemma 2.2), let $h_w = f \circ g \in \mathbb{R}(z)$ be the rational function constructed there. The case $w = 0_E$ gives a Lattès function $h_0 \in \mathbb{R}(z)$. It is easy to see that $h_w = \lambda_1 \circ h_0 \circ \lambda_2$ for linear fractional functions $\lambda_1, \lambda_2 \in \mathbb{C}(z)$. If w has the property from Lemma 2.2, then λ_1, λ_2 cannot be chosen in $\mathbb{R}(z)$. So h_w is a twist of h_0 over a quadratic field. Therefore, the relation of h_w to the Lattès map h_0 is analogous to the relation of Rédei functions to cyclic polynomials z^n .

A construction like h_w appeared in an arithmetic context in [7].

Lattès functions, which were known before Lattès work in 1918, are classical objects in complex analysis. See [12] and [9] for the relevance of these functions in complex dynamics, and especially [9] for a lot of information about the history of these functions. Lattès functions also appeared in 1877 in the context of approximation theory in work by Zolotarev. Today they are called Zolotarev functions in approximation theory.

Example 2.4. Here we explicitly compute an example for the case $\ell = 3$. We aim to find an example where the elliptic curve E is defined over \mathbb{Q} , $f \circ g \in \mathbb{Q}(z)$, and $f, g \in K(z)$, where K is an as small as possible number field. Let ω be a primitive third root of unity, so $\omega^2 + \omega + 1 = 0$ and $\bar{\omega} = -1 - \omega$. As $c \in E(\mathbb{C})$ is required to be a non-real point, and the coordinates of the ℓ -torsion group of an elliptic curve over \mathbb{Q} generate the field of ℓ -th roots of unity, we necessarily have $\omega \in K$. Indeed, there are examples with $K = \mathbb{Q}(\omega)$.

The in terms of the conductor smallest elliptic curve E over \mathbb{Q} which has a 3-torsion point in $E(K) \setminus E(\mathbb{Q})$ has the Cremona label 14a2 and Weierstrass form $Y^2 = X^3 - 46035X - 3116178$. One computes that $c = (72\omega - 33, 1080\omega - 648) \in E(\mathbb{C})$ has order 3. Set $C = \langle c \rangle$. Then $C \cap \bar{C} = \{0_E\}$. There is an isogeny $\Phi : E \rightarrow E'$ with kernel C , where E' is given by $Y^2 = X^3 + (298080\omega + 537165)X + (86819040\omega - 39204594)$.

Set $w = (-78, 0) \in E(\mathbb{Q})$. The X -coordinates of \hat{w} with $2\hat{w} = w$ are roots of $X^2 + 156X + 33867 = (X + 78)^2 + 27783$, so there is no $\hat{w} \in E(\mathbb{R})$ with $2\hat{w} = w$.

Thus E, C and w fulfill all the assumptions which we needed in the existence proof of $f(z)$ and $g(z)$. We now compute these functions. Let β and β' be the automorphisms of E and E' given by $\beta(p) = w - p$ and $\beta'(p') = \Phi(w) - p'$. Write $w = (w_x, w_y)$ and $\Phi(w) = (w'_x, w'_y)$. Set $z = \frac{w_y + y}{w_x - x}$ and $z' = \frac{w'_y + y'}{w'_x - x'}$. Recall that $\Phi^*(x') = A(x)$, where $\Phi((u, v)) = (A(u), B(u)v)$ for all $(u, v) \in E(\mathbb{C})$. From that we see also $\Phi^*(y') = B(x)y$.

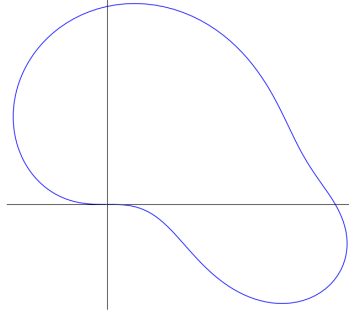
Now recall that the function $g(z)$ we are looking for fulfills $g(z) = \Phi^*(z')$. We compute

$$g(z) = \Phi^* \left(\frac{w'_y + y'}{w'_x - x'} \right) = \frac{w'_y + B(x)y}{w'_x - A(x)}.$$

Use this equation, and the equations $z = \frac{w_y + y}{w_x - x}$ and $y^2 = x^3 - 46035x - 3116178$ to eliminate the variables x and y . So we are left with a polynomial equation in z and the unknown function $g(z)$ which we treat as a variable. This polynomial has a factor of degree 1 with respect to $g(z)$, from which we obtain $g(z)$. Analogously we get $f(z)$. After minor linear changes over \mathbb{Q} (which slightly simplify f and g) we obtain

$$\begin{aligned} f(z) &= \frac{z^3 - 6(\omega + 1)z}{3z^2 + 1} \\ g(z) &= \frac{2z^3 + (\omega + 1)z}{z^2 - \omega} \\ f(g(z)) &= \frac{8z^9 - 24z^5 - 13z^3 - 6z}{12z^8 + 13z^6 + 12z^4 - 1} \in \mathbb{Q}(z). \end{aligned}$$

The following plot shows the image of $\hat{\mathbb{R}}$ under $\frac{1}{1+g(z)}$. As expected, this curve is a



Jordan curve, but not a circle.

3 Proof of Proposition 1.6

If G acts on a set Ω , then ω^g denotes the image of $\omega \in \Omega$ under $g \in G$. Furthermore, $G_\omega = \{g \in G \mid \omega^g = \omega\}$ is the stabilizer of ω in G .

For $g, h \in G$ we write g^h for the conjugate $h^{-1}gh$ of g under h . Similarly, if S is a subset or subgroup of G , then $S^h = \{s^h \mid s \in S\}$.

If G is transitive on Ω , then $\emptyset \neq \Delta \subseteq \Omega$ is called a block if $\Delta = \Delta^g$ or $\Delta \cap \Delta^g = \emptyset$ for each $g \in G$. If this is the case, then Ω is a disjoint union of sets $\Delta_i = \Delta^{g_i}$ for g_i in a subset of G . These sets Δ_i are called a block system. Note that G acts by permuting these sets Δ_i .

We assume that Proposition 1.6 is false, so there is a group M with $G_\omega \leq M \leq G$ and $M \neq M^\sigma$. Among the counterexamples with $|G|$ minimal we pick one with $|\Omega|$ minimal. In a series of lemmas we derive properties of such a potential counterexample, and eventually we will see that it does not exist.

Note that G_ω^σ fixes $\omega^\sigma = \omega$, hence $G_\omega^\sigma \leq G_\omega$ and therefore $G_\omega = G_\omega^\sigma$, a fact we will use frequently. Another trivial fact which we use throughout the proof is the following: If B is a subgroup of G , then σ normalizes $B \cap B^\sigma$ and $\langle B, B^\sigma \rangle$.

Lemma 3.1. *G is transitive on Ω .*

Proof. Set $\Delta = \omega^G$. If $\Delta = \Omega$ then we are done. So assume that $\Delta \subsetneq \Omega$. If $\Delta = \{\omega\}$, then $G_\omega = G$, and therefore of course $M = G = M^\sigma$.

Thus $\{\omega\} \subsetneq \Delta \subsetneq \Omega$. Note that $\Delta^\sigma = \omega^{G^\sigma} = \omega^{\sigma G} = \omega^G = \Delta$. By the assumption of a minimal counterexample, we obtain that the proposition holds for the action of $\langle G, \sigma \rangle$ on Δ , hence $M = M^\sigma$, a contradiction. \square

Lemma 3.2. $G = \langle M, M^\sigma \rangle$.

Proof. Set $H = \langle M, M^\sigma \rangle$. Then $H^\sigma = H$, so if H is a proper subgroup of G , then $M = M^\sigma$ by the minimality assumption of a counterexample. \square

Lemma 3.3. $M \cap M^\sigma = G_\omega$.

Proof. Set $W = M \cap M^\sigma$. Note that $G_\omega \leq W$ and $W^\sigma = W$. Therefore $\Delta = \omega^W$ is a block for the action of $\langle G, \sigma \rangle$ on Ω , and $\Delta^\sigma = \Delta$. Let $\bar{\Omega}$ be the block system which contains Δ , so $\langle G, \sigma \rangle$ acts on $\bar{\Omega}$. By the transitivity of G all blocks in $\bar{\Omega}$ have the same size, and this size divides the odd number $|\Omega|$. So the blocks have odd size, therefore σ has a fixed point in each block which is fixed setwise. Thus Δ is the only block fixed by σ . For $g \in G$ let \bar{g} be the induced permutation on $\bar{\Omega}$. The stabilizer of Δ in \bar{G} is \bar{W} .

Now suppose that $W > G_\omega$, hence $|\Delta| > 1$ and therefore $|\bar{\Omega}| < |\Omega|$. Note that $\bar{W} \leq \bar{M}$. So the proposition applies and yields $\bar{M}^\sigma = \bar{M}$. But the kernel of the map $g \mapsto \bar{g}$ is contained in $W = M \cap M^\sigma$, so $M^\sigma = M$, a contradiction. \square

Lemma 3.4. If $B \leq M$, then either $G = \langle B, B^\sigma \rangle$, or $B \leq G_\omega$.

Proof. Set $H = \langle B, B^\sigma \rangle$, and suppose that $H < G$. Hence the proposition holds for H , in particular $\langle B, H_\omega \rangle^\sigma = \langle B, H_\omega \rangle$. Thus $B \leq \langle B, H_\omega \rangle^\sigma \leq \langle M, G_\omega \rangle^\sigma = M^\sigma$. Together with the previous Lemma we get $B \leq M \cap M^\sigma = G_\omega$. \square

Lemma 3.5. G has even order.

Proof. Suppose that the order of G is odd. Pick $g \in M \setminus G_\omega$. Note that

$$(gg^{-\sigma})^\sigma = g^\sigma g^{-1} = (gg^{-\sigma})^{-1},$$

so σ acts on $\langle gg^{-\sigma} \rangle$ by inverting the elements. As $\langle gg^{-\sigma} \rangle$ has odd order, there is $h \in \langle gg^{-\sigma} \rangle$ with $gg^{-\sigma} = h^2$. Set $c = hg^\sigma$. First note that c is fixed under σ :

$$c^\sigma = (hg^\sigma)^\sigma = h^\sigma g = h^{-1}g = h^{-1}h^2g^\sigma = hg^\sigma = c$$

From this we obtain that c permutes the fixed points of σ , so $c \in G_\omega$ because ω is the only fixed point of σ .

Another calculation shows

$$cg^{-\sigma}c = hg^\sigma g^{-\sigma} hg^\sigma = h^2g^\sigma = g,$$

hence

$$g \in \langle G_\omega, g^{-\sigma} \rangle \leq M^\sigma,$$

contrary to $M \cap M^\sigma = G_\omega$ and the choice of g . \square

Lemma 3.6. *M contains at least one involution which is not contained in G_ω .*

Proof. As $[G : G_\omega] = |\Omega|$ is odd, there is a Sylow 2-subgroup S of G contained in G_ω . By the previous lemma, $|S| > 1$. As the number of Sylow 2-subgroups of G_ω is odd and $G_\omega = G_\omega^\sigma$, we may and do pick S with $S^\sigma = S$.

Let S_2 be the set of involutions in S . Then $S_2^\sigma = S_2$. Suppose that the lemma is false. Then $S_2^m \subseteq G_\omega$ for all $m \in M$. As $\langle S_2^m \rangle = \langle S_2 \rangle^m$ is a 2-group in G_ω , there is $u \in G_\omega$ such that $S_2^{mu} \subseteq S$, and therefore $S_2^{mu} = S_2$. So for each $m \in M$ there is $u \in G_\omega$ such that $mu \in N_M(S_2)$, where $N_M(S_2)$ denotes the normalizer of S_2 in M . Therefore

$$M = \langle N_M(S_2), G_\omega \rangle. \quad (2)$$

From $S_2^\sigma = S_2$ we obtain

$$N_M(S_2)^\sigma = N_{M^\sigma}(S_2). \quad (3)$$

Set $H = \langle N_M(S_2), N_M(S_2)^\sigma \rangle$. If $H < G$, then $N_M(S_2) \leq G_\omega$ by Lemma 3.4, and therefore $M \leq G_\omega$ by (2), a contradiction.

Therefore $H = G$. Together with (3) we obtain

$$G = \langle N_M(S_2), N_M(S_2)^\sigma \rangle = \langle N_M(S_2), N_{M^\sigma}(S_2) \rangle,$$

so all of G normalizes S_2 . Then $Q = \langle S_2 \rangle$ is a nontrivial normal subgroup of G with $Q \subseteq G_\omega$, so Q fixes every point in Ω , a contradiction. \square

We now obtain the final contradiction: Let $a \in M \setminus G_\omega$ be an involution. Set $b = a^\sigma \in M^\sigma$ and let D be the dihedral group generated by a and b . From Lemma 3.4, with $B = \langle a \rangle$, we get $D = G$.

Set $C = \langle ab \rangle$. Then $[G : C] = 2$ (because $G = C \cup Ca$). We claim that C is transitive on Ω . If this were not the case, then, by the transitivity of G and $C \triangleleft G$, C would have exactly two orbits of equal size, so $|\Omega|$ were even, a contradiction.

So $G = CG_\omega$, and $C \cap G_\omega = 1$, because transitive abelian groups act regularly. The modular law yields $M = (C \cap M)G_\omega$ and $M^\sigma = (C \cap M^\sigma)G_\omega$.

From $|M| = |M^\sigma|$ we get $|C \cap M| = |C \cap M^\sigma|$. But the subgroups of the cyclic group C are determined uniquely by their order, hence $C \cap M = C \cap M^\sigma$ and finally $M = M^\sigma$.

4 Proof of Theorems 1.4 and 1.5

For the rational function $g(z) \in \mathbb{C}(z)$ let $\bar{g}(z)$ be the function with complex conjugate coefficients. Recall that $g(\mathbb{R})$ is a circle in $\hat{\mathbb{C}}$ if and only if there is a linear fractional function $\lambda \in \mathbb{C}(z)$ such that $\lambda \circ g \in \mathbb{R}(z)$. The following lemma gives a useful necessary and sufficient criterion for this to hold. By $\mathbb{C}(g(z))$ we mean the field of rational functions in $g(z)$.

Lemma 4.1. *Let $g(z) \in \mathbb{C}(z)$. Then $\lambda \circ g \in \mathbb{R}(z)$ for some linear fractional function $\lambda \in \mathbb{C}(z)$ if and only if $\mathbb{C}(g(z)) = \mathbb{C}(\bar{g}(z))$.*

Proof. If $\lambda \circ g \in \mathbb{R}(z)$, then $\lambda \circ g = \overline{\lambda \circ g} = \bar{\lambda} \circ \bar{g}$, hence $\bar{g}(z) = \bar{\lambda}^{-1}(\lambda(g(z)))$, and therefore $\mathbb{C}(\bar{g}(z)) = \mathbb{C}(g(z))$.

To prove the other direction, suppose that $\mathbb{C}(g(z)) = \mathbb{C}(\bar{g}(z))$. This assumption is preserved upon replacing g with $\mu \circ g$ for a linear fractional function $\mu \in \mathbb{C}(z)$. Thus, without loss of generality, we may assume that there are $r_1, r_2, r_3 \in \mathbb{R}$ with $g(r_1) = \infty$, $g(r_2) = 0$, $g(r_3) = 1$. From $\mathbb{C}(g(z)) = \mathbb{C}(\bar{g}(z))$ we get $\bar{g} = \rho \circ g$ for a linear fractional function $\rho \in \mathbb{C}(z)$. Evaluating in r_1, r_2 , and r_3 yields that ρ fixes $\infty, 0$ and 1 , hence $\rho(z) = z$. So $\bar{g} = g$, and therefore $g \in \mathbb{R}(z)$. \square

Remark 4.2. The lemma holds more generally if we replace \mathbb{R} with a field K and \mathbb{C} with a Galois extension E of K , and $\mathbb{C}(g(z)) = \mathbb{C}(\bar{g}(z))$ by the condition $E(g(z)) = E(g^\sigma(z))$ for all $\sigma \in \text{Gal}(E/K)$. Indeed, if K is an infinite field, then we find $r_1, r_2, r_3 \in K$ such that the values $g(r_1), g(r_2)$, and $g(r_3)$ are distinct and therefore without loss of generality equal to $\infty, 0$ and 1 . So, as above, $g = g^\sigma$ for all $\sigma \in \text{Gal}(E/K)$. Thus the coefficients of g are fixed under $\text{Gal}(E/K)$ and therefore contained in K .

If K is finite, we can argue as follows: We may assume that $g(\infty) = \infty$, so $g(z) = p(z)/q(z)$ for relatively prime polynomials $p, q \in E[z]$ with $\deg p > \deg q$. In addition, we may assume that p and q are monic. Let σ be a generator of the cyclic group $\text{Gal}(E/K)$. From $g^\sigma(z) \in E(g(z))$ and $g^\sigma(z) = \frac{p^\sigma(z)}{q^\sigma(z)}$ we obtain $g^\sigma = g + b$ for some $b \in E$. Repeated application of σ shows that $\text{Trace}_{E/K} b = 0$. So by the additive Hilbert's Theorem 90 there is $c \in E$ with $c - c^\sigma = b$, hence $(g + c)^\sigma = g + c$ and therefore $g + c \in K(z)$. (The same argument, except that Hilbert's Theorem 90 is a trivial fact for the extension \mathbb{C}/\mathbb{R} , works as an alternative proof of the lemma too.)

Theorem 1.4 is a direct consequence of Theorem 1.5. For if g is weakly injective, and $g \circ f$ is injective on $g(\hat{\mathbb{R}})$, then $g \circ f \circ g$ is weakly injective, so $f \circ g$ is weakly injective even more.

Thus we only need to prove Theorem 1.5.

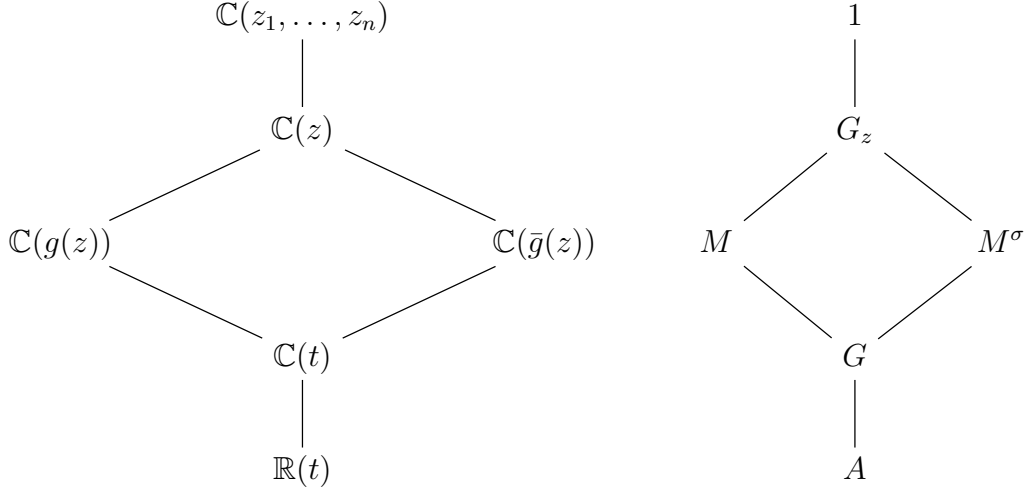
Let t be a variable over \mathbb{C} , and Z be another variable over the field $\mathbb{C}(t)$ of rational functions in t .

If $h(\infty) \neq \infty$, then upon replacing $h(Z)$ with $\frac{1}{h(Z) - h(\infty)}$ (and τ with $\frac{1}{\tau - h(\infty)}$) we may assume that $h(\infty) = \infty$. By further replacing $h(Z)$ with $h(Z) - \tau$, we may and do assume that $h(Z) = \frac{p(Z)}{q(Z)}$, where $p(Z), q(Z) \in \mathbb{R}[Z]$ are relatively prime polynomials, $\deg p(Z) = n > \deg q(Z)$, and $p(Z) = \prod_{i=1}^n (Z - \alpha_i)$, where the α_i are pairwise distinct, $\alpha_1 \in \mathbb{R}$, and $\alpha_i \notin \mathbb{C} \setminus \mathbb{R}$ for $i \geq 2$.

By Hensel's Lemma, $p(Z) - tq(Z) = \prod_{i=1}^n (Z - z_i)$, where $z_i \in \mathbb{C}[[t]]$ has constant term α_i . As $\alpha_1 \in \mathbb{R}$ and $p, q \in \mathbb{R}[Z]$, we actually have $z_1 \in \mathbb{R}[[t]]$. Write $z = z_1$. The complex conjugation acts on the coefficients of the formal Laurent series $\mathbb{C}((t))$ and fixes t . Under this action, z is fixed, and the z_i 's for $i \geq 2$ are flipped in pairs. Note that $t = h(z_i)$ for all i .

$\mathbb{C}(z_1, z_2, \dots, z_n)$ is a Galois extension of $\mathbb{R}(t)$, and $\mathbb{C}(z_1, z_2, \dots, z_n) \subseteq \mathbb{C}((t))$. So the restriction of the complex conjugation action of $\mathbb{C}((t))$ to $\mathbb{C}(z_1, z_2, \dots, z_n)$ is an involution $\sigma \in A := \text{Gal}(\mathbb{C}(z_1, z_2, \dots, z_n)/\mathbb{R}(t))$ which fixes $z = z_1$, and moves all z_i with $i > 1$.

Now write $h(z) = f(g(z))$ as in Theorem 1.5. Then also $t = h(z) = f(g(z)) = \bar{f}(\bar{g}(z))$. This yields the following inclusion of fields and the corresponding subgroups of A by Galois correspondence:



Here G_z is the stabilizer of $z = z_1$ in G . As M is the stabilizer of $g(z)$ in G , and σ maps $g(z)$ to $\bar{g}(z)$, the stabilizer of $\bar{g}(z)$ is $\sigma^{-1}M\sigma = M^\sigma$.

By construction, $\mathbb{C}(z_1, z_2, \dots, z_n)$ is the splitting field of $p(Z) - tq(Z)$ over $\mathbb{C}(t)$, hence G acts faithfully on $\{z = z_1, z_2, \dots, z_n\}$. Now G is normal in A , so $\sigma \in A$ normalizes G . Furthermore, σ fixes exactly one of the z_i . So $M = M^\sigma$ by Proposition 1.6. Thus $\mathbb{C}(g(z)) = \mathbb{C}(\bar{g}(z))$ by the Galois correspondence, and finally $\lambda(g(z)) \in \mathbb{R}(z)$ for some linear fractional $\lambda \in \mathbb{C}(z)$ by Lemma 4.1. This proves Theorem 1.5.

5 Some more examples

If $h = f \circ g$ for polynomials $f, g \in \mathbb{C}[z]$, and $h \in \mathbb{R}[z]$, then it is well known that there is a linear polynomial $\lambda \in \mathbb{C}[z]$ such that $\lambda \circ g \in \mathbb{R}[z]$. See [6, Theorem 3.5], or [13, Prop. 2.2] for a down to earth proof. A less elementary but more conceptual proof can be based on the fact that the Galois group of $h(Z) - t$ over $\mathbb{C}(t)$ contains an element which cyclically permutes the roots of $h(Z) - t$, and the other fact that subgroups of cyclic groups are uniquely determined by their orders.

Note that if $h = f \circ g$ for a polynomial h and rational functions f, g , then there is a linear fractional function $\rho \in \mathbb{C}(z)$ such that $h = (f \circ \rho^{-1}) \circ (\rho \circ g)$, and $f \circ \rho^{-1}$ and $\rho \circ g$ are polynomials. (This follows from looking at the fiber $h^{-1}(\infty)$.)

So in order to get examples of rational functions $h \in \mathbb{R}(z)$ which decompose as $h = f \circ g$ with $f, g \in \mathbb{C}(z)$ such that there is no linear fractional function $\lambda \in \mathbb{C}(z)$ with $\lambda \circ g \in \mathbb{R}(z)$, one has to assume that h is not a polynomial.

One also has to assume that g is not a polynomial, as the following easy result shows.

Lemma 5.1. *Suppose that $f \circ g \in \mathbb{R}(z)$ where $f \in \mathbb{C}(z)$ and $g \in \mathbb{C}[z]$ are not constant. Then $\lambda \circ g \in \mathbb{R}[z]$ for a linear polynomial $\lambda \in \mathbb{C}[z]$.*

Proof. Assume without loss of generality that g is monic. Write $f = \frac{p}{q}$ with $p, q \in \mathbb{C}[z]$ relatively prime and p monic. From $f \circ g \in \mathbb{R}(z)$ we obtain

$$\frac{\bar{p}(\bar{g}(z))}{\bar{q}(\bar{g}(z))} = \frac{p(g(z))}{q(g(z))}.$$

Clearly, both fractions are reduced, and the numerators of both sides are monic. Therefore $\bar{p}(\bar{g}(z)) = p(g(z))$, hence $p \circ g \in \mathbb{R}[z]$, and the claim follows from the polynomial case. \square

Now we give some examples of rational functions $h \in \mathbb{R}(z)$ with a decomposition $h = f \circ g$ with $f, g \in \mathbb{C}(z)$ which is not equivalent to a decomposition over \mathbb{R} . Recall that this is equivalent to $g(\hat{\mathbb{R}})$ not being a circle. Of course, Theorem 1.2 gives many example for this. But these examples are quite complicated and not explicit. However, if one drops the requirement that $g(\hat{\mathbb{R}})$ is a Jordan curve, then there are quite simple examples. We give two series.

Example 5.2 (Attributed to Pakovich by Eremenko in [3]). Let $T_n \in \mathbb{R}[z]$ be the polynomial with $T_n(z + \frac{1}{z}) = z^n + \frac{1}{z^n}$ (so T_n is essentially a Chebychev polynomial.) Set $g(z) = \zeta z + \frac{1}{\zeta z}$ for an n -th root of unity ζ . Then $T_n(g(z)) = z^n + \frac{1}{z^n} \in \mathbb{R}(z)$, while $g(\hat{\mathbb{R}})$ is not a circle if $\zeta^4 \neq 1$.

Example 5.3. Pick $\zeta \in \mathbb{C}$ with $|\zeta| = 1$, and set

$$F = z^k(1 - z)^{n-k}, \quad G = \frac{1 - \zeta z^k}{1 - \zeta z^n}, \quad \mu(z) = \frac{z + i}{z - i}$$

for $1 \leq k < n$. A straightforward calculation shows that

$$F(\bar{G}(\frac{1}{z})) = \frac{1}{\zeta^{n-k}} F(G(z)). \quad (4)$$

Pick $\rho \in \mathbb{C}$ with $\rho^2 = \frac{1}{\zeta^{n-k}}$, and set

$$f(z) = \rho F(z), \quad g(z) = G(\mu(z)).$$

From $\bar{\mu}(z) = \mu(\frac{1}{z})$ and (4) we get $\overline{f \circ g} = f \circ g$, hence $f \circ g \in \mathbb{R}(z)$.

On the other hand, it is easy to see that, except for a some degenerate cases, $g(\hat{\mathbb{R}})$ is not a circle. Furthermore, we see that $g(\hat{\mathbb{R}})$ isn't even a Jordan curve (unless it is a circle), for if z runs through $\hat{\mathbb{R}}$, $\mu(z)$ runs through the unit circle, so the numerator and denominator of $g(z) = G(\mu(z))$ vanish k and n times, respectively, so $g(\hat{\mathbb{R}})$ has several self intersections.

Originally I had only found the cases $\zeta = -1$, $k = n - 1$. Mike Zieve observed the strong similarity of these examples with functions which turned up in work of Avanzi and Zannier. In [1] they classify triples $F \in \mathbb{C}[z]$, $G_1, G_2 \in \mathbb{C}(z)$ such that $F \circ G_1 = F \circ G_2$. One of their cases ([1, Prop. 4.7(3)]) is the above series with $\zeta = 1$, and the series [1, Prop. 5.6(4)] is essentially our series from above.

The connection with the work by Avanzi and Zannier is not a surprise: If we look for polynomials $F \in \mathbb{R}[z]$ such that there is $G \in \mathbb{C}(z) \setminus \mathbb{R}(z)$ with $F \circ G \in \mathbb{R}(z)$, then $F \circ \bar{G} = F \circ G$ with $G \neq \bar{G}$. Furthermore, note that if ζ is an m -th root of unity, then $f(z)^{2m} \in \mathbb{R}(z)$. So upon setting $\tilde{f} = f^{2m} = F^{2m}$, we have $\tilde{f} \circ g = \tilde{f} \circ \bar{g}$.

References

- [1] Roberto M. Avanzi and Umberto M. Zannier. The equation $f(X) = f(Y)$ in rational functions $X = X(t)$, $Y = Y(t)$. *Compositio Math.*, 139(3):263–295, 2003.
- [2] Alexandre Eremenko. Circles and rational functions. MathOverflow. URL:<http://mathoverflow.net/q/103949> (version: 2012-12-03).
- [3] Alexandre Eremenko. Invariant curves and semiconjugacies of rational functions. *Fund. Math.*, 219(3):263–270, 2012.
- [4] Alexandre Eremenko. Analytic invariant curves for rational functions, 2013. Unpublished note.
- [5] Alexandre Eremenko. Personal communication, 2013.
- [6] Michael D. Fried and Robert E. MacRae. On the invariance of chains of fields. *Illinois J. Math.*, 13:165–171, 1969.
- [7] Robert M. Guralnick, Peter Müller, and Jan Saxl. *The Rational Function Analogue of a Question of Schur and Exceptionality of Permutation Representations*, volume 162 of *Memoirs of the American Mathematical Society*. American Mathematical Society, 2003.
- [8] Robert E. MacRae and Pierre Samuel. Subfields of index 2 of elliptic function fields. In *Conference on Commutative Algebra (Univ. Kansas, Lawrence, Kan., 1972)*, pages 171–193. Lecture Notes in Math., Vol. 311. Springer, Berlin, 1973.
- [9] John Milnor. On Lattès maps. In *Dynamics on the Riemann sphere*, pages 9–43. Eur. Math. Soc., Zürich, 2006.
- [10] Peter Müller. Circles and rational functions. MathOverflow. URL:<http://mathoverflow.net/q/115979> (version: 2012-12-11).
- [11] Joseph H. Silverman. *Advanced Topics in the Arithmetic of Elliptic Curves*. Springer-Verlag, New York, 1994.
- [12] Joseph H. Silverman. *The Arithmetic of Dynamical Systems*, volume 241 of *Graduate Texts in Mathematics*. Springer, New York, 2007.
- [13] Gerhard Turnwald. On Schur’s conjecture. *J. Austral. Math. Soc. Ser. A*, 58:312–357, 1995.

INSTITUT FÜR MATHEMATIK, UNIVERSITÄT WÜRZBURG, 97074 WÜRZBURG, GERMANY

E-mail: `peter.mueller@mathematik.uni-wuerzburg.de`